2018-11-15

Spatial and temporal patterns of Pseudo-nitzschia genetic diversity in the North Pacific Ocean from the Continuous Plankton Recorder survey

Stern, R

10.3354/meps12711
Marine Ecology Progress Series
Inter-Research Science Center

Spatial and temporal patterns of *Pseudo-nitzschia* genetic diversity in the North Pacific Ocean from Continuous Plankton Recorder surveys

Rowena Stern†*[1], Stephanie K. Moore†[2], Vera L. Trainer[3], Brian D. Bill[3], Astrid Fischer[1], Sonia Batten[4]

[1]Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, Plymouth, PL1 2PB, UK

[2]University Corporation for Atmospheric Research, Joint Office for Science Support. Visiting Scientist at Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112 USA

[3]Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E., Seattle, WA 98112 USA

[4]Sir Alister Hardy Foundation for Ocean Science, C/O 4737 Vista View Cr, Nanaimo, BC V9V 1N8, Canada

† joint first authors

*Corresponding author e-mail: rost@sahfos.ac.uk

Keywords: *Pseudo-nitzschia*, CPR, Pacific Ocean, PDO, NGS

Abstract

Several species of the marine diatom *Pseudo-nitzschia* can produce the neurotoxin domoic acid that is responsible for the seafood-borne illness amnesic shellfish poisoning in humans, marine wildlife mortalities, and prolonged closures of fisheries resulting in economic losses to coastal communities. Since the year 2000, *Pseudo-nitzschia* species have been monitored in the Pacific Ocean with the Continuous Plankton Recorder (CPR). This study used a combination of scanning electron microscopy with high-throughput and Sanger sequencing of CPR survey samples to compare the diversity of phytoplankton, including *Pseudo-nitzschia* species, from the north-eastern Pacific Ocean over three climatically different years: 2002, 2005, and 2008. Using a *Pseudo-nitzschia*-specific primer set targeting a 320bp region of the large subunit ribosomal DNA (rDNA), revealed spatially-separated communities of *Pseudo-nitszschia*. The coastal region was dominated by a diverse array of *Pseudo-nitzschia fraudulenta* unique sequences (OTUs) whilst the offshore region was rich in *P. multiseries* along with and contained a wide range of other *Pseudo-nitzschia* taxa, many not observed in this region. In 2008, exceptionally cold sea surface temperatures were observed, influenced

by a strong negative Pacific Decadal Oscillation signal. In that year, a more diverse assemblage of species was present in a Spring open water sample whilst *P. fraudulenta* was unusually rare in a coastal Autumn sample. This is the first application of high-throughput genetic methods to uncover patterns of *Pseudo-nitzschia* genetic diversity from archival CPR samples, demonstrating the value of using CPR for plankton community analysis in rarely sampled regions of the oceans.

1. Introduction

Marine diatoms in the genus *Pseudo-nitzschia* are closely monitored in the eastern Pacific Ocean due to their capacity to produce the potent neurotoxin domoic acid (DA). DA can accumulate in filter-feeding fish and shellfish and be transferred through foodwebs to poison humans, marine mammals and seabirds (Work et al. 1993, Scholin et al. 2000). Symptoms of this poisoning in humans, called amnesic shellfish poisoning (ASP), include gastrointestinal distress, seizures, coma, and permanent short-term memory loss, with severe intoxications resulting in death (Perl et al. 1990). Monitoring programs exist worldwide to protect human health from the effects of ASP. For example, in Washington State, USA, regular beach monitoring is conducted to look for cells of *Pseudo-nitzschia* in coastal waters (Trainer & Suddleson 2005) and shellfish are regularly tested for DA by the Washington State Department of Health. Shellfish harvesting closures are implemented when concentrations of DA exceed the regulatory limit for human consumption of 20 ppm in shellfish meat tissue. The first closure of recreational and commercial shellfish harvesting due to DA on the Washington State coast occurred in 1991 and resulted in a $15-20 million revenue loss to local fishing communities (Horner & Postel 1993, Anderson 1995). The total estimated economic impact associated with a coastwide, year-long closure of the razor clam fishery, such as those that occurred in 1991-1992, 1998-1999, and 2002-2003, has been estimated at $21.9 million (Dyson & Huppert 2010).

The Pacific Decadal Oscillation (PDO) is a pattern of ocean-climate variability that gives rise to very different climate regimes with implications for environmental parameters that influence *Pseudo-nitzschia* growth and toxicity. The PDO index is the first mode of monthly ocean sea surface temperature (SST) variability in the North Pacific Ocean poleward of 20°N (Mantua et al. 1997). When the PDO index is positive (negative), the coastal ocean in the Pacific Northwest is typically warmer (cooler) and the central north Pacific Ocean is cooler (warmer) (Mantua et al. 1997). The regional climate is also influenced by the PDO, with

winter-time air temperature and precipitation in the USA Pacific Northwest typically below normal during warm phases of the PDO. Historically, the warm and cool phases of the PDO have persisted for 20-30 years, but in recent years the PDO has been switching phases approximately every 5 years and has closely tracked the El Niño/Southern Oscillation (ENSO). The mechanisms that give rise to the PDO are not fully understood; nevertheless, major changes in marine ecosystems and the distribution and ratios of nutrients in the Pacific Ocean have been documented to occur when the PDO changes phase (Botsford et al. 1997, Mantua et al. 1997). In general, biological productivity is enhanced off the coast of Alaska and inhibited off the coast of the contiguous USA during warm phases of the PDO, while the reverse is true during cold phases (Hare 1999). Phytoplankton communities, including *Pseudo-nitzschia* species, may be affected by changing temperature, salinity and nutrient distributions that may co-occur with PDO phase changes. In fact, recent work suggests that warm phases of the PDO (and ENSO) are directly related to USA west coast toxic *Pseudo-nitzschia* bloom events (McCabe et al. 2016).

The Continuous Plankton Recorder (CPR) is an instrument designed to be towed from merchant ships on their normal sailings and provides opportunities for sampling plankton communities in rarely sampled regions of the open oceans. It works by filtering plankton on a moving band of silk mesh over long distances. The CPR survey was originally designed to collect zooplankton and higher abundances of larger phytoplankton. As such, the silk gauze that collects plankton has a mesh size of ~270 µm. Collection of phytoplankton by the CPR survey would be considered suboptimal, yet phytoplankton to 5 µm (coccolithophorids) have been retained (Richardson et al. 2006). This is because the large volume of water filtered (3 m^3) deposits large amounts of plankton that clog the mesh, effectively reducing the aperture size and retaining smaller plankton (Batten et al. 2003b). Additionally, phytoplankton can be trapped on the silk collecting gauze; the silk material is thicker and stickier than nylon used in plankton nets, with micro-threads that extend into the aperture. The CPR measurement of phytoplankton colour index (PCI), a proxy for total phytoplankton abundance, also correlates well with fluorometric and satellite-measured chlorophyll *a*, although seasonally variable (Batten et al. 2003a, Raitsos 2005). *Pseudo-nitzschia* species are typically between 40-175 µm long, smaller than the mesh size, but can occur in long chains and so may be retained more readily than other smaller or non-chain forming phytoplankton.

The CPR survey monitors phytoplankton from ships of opportunity on two routes in the North Pacific (Batten 2006). One is a 3000-km trans-Pacific route from Vancouver, Canada to Hokkaido, Japan, through subpolar waters. This latter route has been sampled seasonally since 2000 during both warm and cool phases of the PDO. CPR samples are immediately preserved in formalin and archived, and offer an opportunity to examine the spatial distribution of *Pseudo-nitzschia* species over different temperature and ocean-climate regimes. At present, *Pseudo-nitzschia* retained on the mesh are examined microscopically and classified into two cell-width morphotypes, *P. seriata* (>3µm) complex and *P. delicatissima* (<3µm) complex (hereafter referred to as *P. seriata* and *P. delicatissima*-sized cells, respectively). Identification to lower taxonomic levels is not possible due to the limitation of light microscopy in identifying the minute morphological differences between species (Hasle 1993). Because of the cryptic and pseudo-cryptic morphological diversity of *Pseudo-nitzschia* species, morphological and genetic taxonomic approaches are now often used in tandem (Lundholm et al. 2006). Most studies use all or part of the ribosomal internal transcribed spacer (ITS) for identification, which has been found to distinguish species and even intraspecific populations within species (Lundholm et al. 2003, Orsini et al. 2004, Amato et al. 2007, Hubbard et al. 2008, Andree et al. 2011, Lim et al. 2012, Penna et al. 2012). The large subunit (LSU) ribosomal DNA (rDNA) has also been used successfully, although with a lesser degree of resolution to species or species groups (Lundholm et al. 2002, McDonald et al. 2007).

The use of genetic taxonomic approaches to identify *Pseudo-nitzschia* species from archived samples can be limited by how the samples are preserved. Despite the use of buffered-formalin to reduce hydrolytic fragmentation of DNA molecules, formalin-preservation still causes methylation as well as methylol modification of nucleobases and cross-linking between nucleotides or together with proteins (Paireder et al. 2013, Karmakar et al. 2015). Therefore, genetic analysis of formalin-preserved CPR samples presents challenges. Nevertheless, recent successes in genetic identification of species from CPR samples dating as far back as 1961 include the coccolithophore *Emiliania huxleyi* (Ripley et al. 2008), various microbial eukaryotes from 1 µm in size (McQuatters-Gollop et al. 2015), the harmful algae *Karenia mikimotoi* (Al-Kandari 2012) and the bacterium *Vibrio cholerae* (Vezzulli et al. 2012, Vezzulli et al. 2016). The use of 454 GS FLX+ high-throughput sequencing technology (HTS), or similar HTS technology such as MiSeq (Illumina) are suitable for environmental barcoding of samples, as it uses small (150-300 bp) amplicon sizes and

provides 500-1000Mb per run (Scholz et al. 2012) .In this study, we examined *Pseudo-nitzschia* species assemblages in the eastern North Pacific Ocean region in oligotrophic open waters compared to coastal waters off Vancouver Island, Canada, during both warm PDO (2002 and 2005) and cool (2008) phases of the PDO. We use rDNA LSU primers designed for the genus *Pseudo-nitzschia* (McDonald et al. 2007) to determine species-distributions in thirty CPR samples. Ten of these samples were able to generate PCR products for HTS and Sanger sequencing of Clone-libraries of PCR products (CLS), providing a species-level comparison of *Pseudo-nitzschia* diversity in coastal and open Pacific waters.

2. Methods

2.1 CPR samples

The CPR is deployed on the trans-Pacific route between Vancouver, Canada and Hokkaido, Japan every 3 months. CPR transects along the route were divided into two regions; (1) the Eastern region, including the shelf of North America to -134˚E plus one sample at -136˚E, and (2) the Central region, including the open ocean region from -134˚ to -148˚E (Fig. 1). Thirty samples out of a total of 159 were initially selected for genetic analysis to represent three seasons (spring, summer and autumn) during 2002, 2005 and 2008 (Fig. 1 and Table 1). Eleven of the thirty samples successfully generated genetic results (see below). Samples were chosen on the basis of high *Pseudo-nitzschia* abundance determined from light microscopy. Mean abundances of total diatoms and *Pseudo-nitzschia* species from all 159 samples were calculated for each season, year, and region, to compare the community composition. In Fig. 7, the mean abundance of diatoms and *Pseudo-nitzschia* cells were calculated for central or eastern Pacific regions per season per year (termed seasonal means) from standard cell counts of all CPR samples (total 159) from 2002, 2005 and 2008 so that there was 2-4 CPR samples per seasonal mean.

2.2 Phytoplankton community analysis

Phytoplankton taxa were identified and counted from CPR samples as described in Batten et al. (2003a). Hard-shelled phytoplankton were counted under a light microscope by viewing 20 fields of view (diameter 295 μm) across each sample under high magnification (× 450) and recording the presence of all the taxa in each field (presence in 20 fields is assumed to reflect a more abundant organism). These 20 fields amount to 1/10,000 of the area of the

filtering silk. Cell abundances per field (H) were then calculated for each taxon (Robinson & Hiby 1978):

$$H = -\ln(k/20) \quad (1)$$

where k is the number of empty microscope fields (out of 20) observed. Multiplication by the proportion of the sample examined gives cell abundances in each sample. A category system is used to calculate the average abundance per sample, ranging from 0-750,000 per sample (for a full explanation of the sampling technique see (Richardson et al. 2006)). The two main groups of *Pseudo-nitzschia* that are routinely recorded in CPR samples are distinguished by their width in valve view, with the *Pseudo-nitzschia delicatissima* sized cells being smaller than 3 µm in width and the *P. seriata*-sized cells having a width exceeding 3 µm. Inconclusive species are recorded as *Nitzschia* spp. The mean sample taxonomic abundances for each year/region/season unit were transformed using $\log^{10}(x+1)$, where x is abundance, for all 159 CPR samples.

2.3 DNA extraction

Each CPR sample represents a collection over 10 nautical miles and is equivalent to filtering ~3 m^3 of water (Richardson et al. 2006). A quarter piece of a CPR sample was cut so that it represented the entire 10 nautical miles but only a quarter of the volume of filtered plankton (0.75 m^3). The CPR silk piece was cut into 1-cm^2 square pieces and placed into 30 mL of TE buffer. The procedure for extracting DNA is described in detail elsewhere (Ripley et al. 2008) and is only briefly described here. The CPR silk piece was washed and agitated in TE buffer for 24 hours, the plankton was recovered by centrifugation, resuspended into 1 mL fresh TE buffer and divided in two 500 µL duplicate samples. Both duplicates were treated with Proteinase K and sodium dodecyl sulphate (SDS) for 48 hours, followed by a phenol/chloroform/isoamyl alcohol (25:24:1) extraction. The upper aqueous layer from the phenol-chloroform step was further extracted by chloroform/isoamyl alcohol (24:1). DNA was precipitated with ammonium chloride and ethanol extraction and the DNA was resuspended in 30 µL of TE buffer.

2.4 PCR amplification and sequencing

PCR amplification on 30 CPR samples and genomic DNA from two non-preserved cultures of *Pseudo-nitzschia multiseries* (culture lost) and *Pseudo-nitzschia fraudulenta* (CCAP1061/6) from the Culture Collection of Algae and Protozoa (SAMS, Scotland) was attempted using a 600-800bp LSU marker (Scholin 1994)and ITS markers (White 1990,

Hubbard et al. 2008, Andree et al. 2011). The ITS marker amplifications yielded no amplicons except for very faint products for samples 139VJ5, 139VJ37 and 146VJ5 and genomic *P. fraudulenta* DNA. Amplification of diluted genomic DNA (1:10, 1:100, 1:1000) in a subset of samples also failed. A number of nested PCR strategies were used for ITS amplification with no success. With most amplification reactions (except for these three samples) resulting in failure using the ITS marker, it was eliminated from this study (see supplementary Table A1). However, a nested PCR amplification approach using LSU markers was successful in yielding products in CPR samples and the cultures. General eukaryotic LSU primers D1R and D2C (Scholin 1994) resulted in 22/30 amplicons from CPR samples (size 600-800 bp). Amplifications were carried out with the Promega PuReTaq kit (Promega, WI, USA) using 2 µL of genomic DNA (ranging from 25-1073 ng/µL, mean 288 ng/µL) which were then diluted by 1:100 in a reaction volume of 25 µL containing 3 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM each of primers, , and 1 unit of Taq polymerase. PCR conditions were 95˚C for 5 minutes, then 35 cycles of denaturation at 95˚C for 30 seconds, annealing at 45˚C for 45 seconds and extension 72˚C for 45 seconds and a final extension step of 72˚C for 10 minutes. A *Pseudo-nitzschia*-specific LSU nested primer set D1-186F and D1-548R (McDonald et al. 2007) was then used on first round PCR products (D1R-D2C) to amplify a 362 bp product that was successful in 10/22 first round amplicons. PCR reaction conditions were as above except 1 µL of first round amplification product was used for a template. PCR cycling conditions were the same as for D1R/D2C, except the annealing temperature was 50˚C and the final 72˚C extension was for 5 minutes.

2.5 Clone library sequence (CLS) analysis

To confirm that only *Pseudo-nitzschia* were amplified using the LSU nested primer set, a clone-library sequencing study was performed on the six nested PCR products that successfully generated positive clones (note that four of the ten nested PCR products failed to generate positive clones). The six samples that successfully generated sequences from clone-libraries are listed in Table 4. The cloning was carried out using the TOPO TA® cloning kit (Life technologies, Paisley, UK) using 25 µl of one Shot® INVαF´ competent cells for 5 µl of PCR product with primer-dimers removed using ExosapIT (Affymetrix, CA, USA). A total of 162 transformed colonies from the remaining samples was prepared for sequencing according to the manufacturer's instructions, except that DNA from colonies were prepared by dissolving a colony into 10 µL of sterile water and heat-denatured at 95˚C for 2 minutes.

Sequencing reactions were performed in 20 µL with 1 µL of BigDye v3.1 and 5× buffer (supplied by Applied Biosystems, CA, USA), 1 µL of 3.2 µM primer (either M13F or M13R) and 20-50 ng of PCR product. The amplicons were sequenced using capillary electrophoresis by Source Bioscience, Nottingham, UK. The CLS dataset was trimmed in using BioEdit v 1999-2013 software (Hall 1999) removing cloning sites, checked using BLASTn (Altschul et al. 1990) for initial identification and added to the HTS dataset (see section 2.7). Repeat sequencing of Sp08C was carried out using freshly re-amplified nested products of D1-186F and D1-548R, as described above, but sequenced using primer PmultLSUR1 (5'GAATCAACCAACCCAAACTCACGCAAGCC 3').

2.6 HTS analysis and OTU generation

To obtain better diversity representation, HTS was conducted on the LSU products of nine samples (listed in Table 4) that contained sufficiently concentrated DNA. Despite a wide range of genomic DNA concentrations, the difference in PCR product concentrations from the *Pseudo-nitzschia* specific nested reaction was no more than 9 ng/µL between samples. All PCR products were diluted to 50 ng/µL and sent to MrDNA Molecular Research Laboratory (Shallowater, Texas, USA) for a custom assay with primers D1-186F and D1-548R, using a single-step 30 cycle PCR using HotStarTaq Plus Master Mix Kit (Qiagen, Valencia, CA) and PCR conditions as described earlier for this primer set. Following the PCR step, amplicon products from all samples were mixed in equal concentrations and purified using Agencourt Ampure beads (Agencourt Bioscience Corporation, MA, USA). Samples were sequenced utilizing Roche 454 FLX titanium instruments and reagents and following manufacturer's guidelines.

Various bioinformatics pipelines incorporated into the Bio-Linux (Field et al. 2006) operating software based on Ubuntu 10.4 were tailored toward the analysis of eukaryotic LSU amplicons. The Python-based QIIME software (Caporaso et al. 2010) script split_libraries.py was used to quality-check reads using default settings and to trim primers and tags. A total of 14906 sequence reads were retrieved from nine samples from HTS sequencing ranging from 2632-6505 reads per sample. Additional filtering criteria were applied with a sliding window quality score of 50 to remove poor quality sequences and to include reads greater than 300 bp (a primer mismatch of one) and manual chimera-checking was performed on aligned sequences (Denoise step). Operational taxonomic unit (OTU) picking steps were performed on denoised sequence data by clustering sequences at 99% and 90% using UCLUST to allow

for abundance pre-sorting (Trobajo et al. 2014) in order to obtain a range of representative taxa. Each OTU is a unique sequence that was at least 1% different to other OTUs. These sequences were exported into BioEdit (Hall 1999) for more precise analysis of OTU identities. Additional quality checks were carried out by BLAST analysis to ensure no chimeras or low quality sequences were retained. All sequences were deposited in Genbank (see supplementary Table A2).

2.7 Phylogenetic analysis of sequences

An initial BLAST search of the 362bp trimmed D1-186F and D1-548R *Pseudo-nitzschia*-specific LSU fragment (McDonald et al. 2007) was carried out to check all HTS and CLS datasets belonged to *Pseudo-nitzschia* and no chimeras were present. Non redundant hits to our sequences that contained species information were used for phylogenetic analysis. We also used the search term "*Pseudo-nitzschia* Large ribosomal" to capture 309 *Pseudo-nitzschia* sequences. An additional 35 other pennate and centric diatom species were added as an outgroup. All reference sequences were downloaded in May 2016 and September 2017. These were combined with environmental (HTS and CLS) and automatically aligned and trimmed using CLUSTALW in BioEDIT (Hall 1999) to 320bp. The alignment was 439 bp long including gaps and contained 768 sequences in total (see supplementary Table A3). The alignment was exported into MEGA 6.0 (Tamura et al. 2013) for phylogenetic analysis using maximum likelihood (ML) method using a Kimura-2 parameter nucleotide substitution model and four Gamma distribution categories to model evolutionary rate differences among sites (4 categories +*G*, parameter = 2.1723). A partial deletion option was selected in which all positions with less than 95% site coverage were eliminated, resulting in 169 positions analysed in the final dataset. ML bootstrap analyses were carried out with 1,000 pseudo-replicates. Initial tree(s) for the heuristic search were obtained by applying the Neighbour-Joining method to a matrix of pairwise distances (PWD) estimated using the Maximum Composite Likelihood (MCL) approach. The tree with the highest log likelihood (-5044.1453) was selected and the percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. Two replicate public sequences were manually removed (JN050300, AF417666). Visualization of the ML tree was only possible by compressing clades that contained a large number of taxa by exporting newick files into interactive tree of life (ITOL (Letunic 2016) and labelled using Adobe Illustrator. An

303 additional ML phylogeny of longer (381bp) CLS reads of the D1-186F, D1-548R LSU
304 fragment from six CPR samples was performed for better identification of environmental
305 sequences (supplementary Fig. A1).The alignment was 430bp long including gaps with 352
306 environmental and public sequences and the phylogeny was and built using the same tree
307 building methods described above (+*G*, parameter = 0.4345) with 280 positions analysed in
308 the final dataset. Investigation of genetic pairwise distances (PWD) was also carried out but
309 did not reveal clear distinction within and between species (see supplementary Fig. A2)
310 Hence PWD metrics were not used to evaluate species here.
311
312 2.8 Scanning electron microscopy
313 In order to confirm the morphological types of *Pseudo-nitzschia* captured by the CPR survey,
314 eight of the genetically analysed CPR samples from the trans-Pacific route from Vancouver,
315 Canada to Hokkaido, Japan during 2002-2008 (see Table 1) and an additional set from 2014
316 were analysed by Scanning Electron Microscopy (SEM) and *Pseudo-nitzschia* cells were
317 identified to species level. Small subsamples of CPR mesh containing preserved
318 phytoplankton material were cut to size, inserted into 15 mL centrifuge tubes and vortexed
319 with 10 mL of MilliQ® water. Subsamples (1-2 mL) were removed and centrifuged in micro-
320 centrifuge tubes. Pellets were rinsed in MilliQ® water 1-2 more times to remove any
321 remaining preservative and then oxidized with 4-5 drops of saturated potassium
322 permanganate solution, cleared with 3 rinses of concentrated hydrochloric acid (HCl) and
323 finally washed in 3 rinses of MilliQ® water to remove HCl. Finally, pellets were resuspended
324 in approximately 0.5 mL MilliQ® water and filtered onto 13 mm diameter, 0.2 µm pore size
325 polycarbonate filters (Millipore Corp.). Filters were then glued to aluminum SEM stubs,
326 coated with gold-palladium and examined in a JEOL 6360LV SEM.
327
328 2.9 Satellite-derived SST and PDO
329 Satellite-derived SST values were obtained on a 1° latitude by 1° longitude grid in the
330 Eastern (-134 to -125°E, 49 to 56°N) and Central (-148 to -134°E, 49 to 56°N) regions of the
331 NE Pacific 3. Optimum Interpolation (OI) SST V2 data are provided by the National
332 Oceanic and Atmospheric Administration (NOAA), Office of Oceanic and Atmospheric
333 Research (OAR), Earth System Research Laboratory (ESRL), Physical Sciences Division
334 (PSD), Boulder, Colorado, USA, from their website at http://www.esrl.noaa.gov/psd/.
335 Seasonal mean values of SST were interpolated for the Eastern and Central regions to
336 determine spatial variability in the regions during seasons and years when CPR samples were

collected. Temporal variability in monthly SST was determined by examining standardized anomalies for grid cells that encompassed locations where CPR samples were collected (grid cells A-H in Fig. 1) from 2000 through 2010. Standardized anomalies were calculated by dividing the anomalies by the climatological standard deviation, using the 11-year baseline period from 2000 through 2010, such that the time series for each grid cell had a mean of zero and a standard deviation of one. Monthly values of the PDO index were obtained from the University of Washington Joint Institute for the Study of the Atmosphere and Ocean (JISAO 2014). Seasonal mean values of the PDO index (sPDO) were calculated for seasons and years when CPR samples were collected.

3. Results

3.1 OTU identification of *Pseudo-nitzschia*

ML phylogenetic analysis of public and environmental sequences using the 362bp *Pseudo-nitzschia*-specific LSU fragment (D1-186F, D1-548R,(McDonald et al. 2007) on 11 CPR samples (Fig.2) identified 28 terminal clades, most with low support. Seventeen of these clades related to single species containing strains identified from previous studies (Table 3, Fig. 2). Seven con-specific clades consisted of two species each (Fig. 2, Table 3).However, due to the lack of resolution of the small region used, these conspecific clades could not be resolved further. *P. galaxiae* and *P. sabit*, identified as sister species by (Teng 2015) split into two sister groups containing different subpopulations of both species (Fig. 2, Table 3). Other species appear in multiple clades, due to the lack of resolution of the smaller LSU fragment which has separated distinct populations such as *P. delicatissima* (Amato et al. 2007, McDonald et al. 2007). *P. brasiliana* was split into a core group and an additional sister clade to *P.americana* and *P.linea*. *P. delicatissima* was found in 5 clades, two of which contained *P. delicatissma* and *P. arenysensis*. A large multi-species clade of *P. delicatissima* clustered with single sequences of *P. turgidula, P. fraudulenta, P. turgidula, P. galaxiae* and *P.pseudodelicatissima*. This larger multi-species group contains a distinct population of *P. delicatissima* (Amato et al. 2007*). P. turgidula,* a common and geographically distinct open Pacific water species, appears twice but no strains could be confirmed to determine the true species group. *P. pseudodelicatissima* appeared in several clades: a core group containing previously identified strains from four confirmed studies including *P. pseudodelicatissima/cuspidata* group (Lundholm et al. 2003, Fernandes 2014) but was indistinguishable from multiple species including other diatom species, *Neodenticula seminae*

371 and *Fragilariopsis* spp. because of a lack of marker resolution. *Pseudo-nitzschia arctica*
372 grouped with one *P. pseudodelicatissima* public sequence from Pacific Northwest (Stehr
373 2002) that may indicate a population of *P. pseudodelicatissima* that is indistinguishable from
374 *P. arctica*, or that these strains are both *P. arctica.* An unknown *Pseudo-nitzschia* sp. genetic
375 clade labelled MVR2015 related to *P. lineola* was also found.
376
377 The remaining 28 other diatom species (excluding *Neodenticula* spp. and *Fragilariopsis* spp.)
378 formed an outgroup that was separate and basal to other *Pseudo-nitzschia* species (Fig. 2).
379 This phylogeny is not as resolved as those using longer LSU reads (Lim et al. 2013) but there
380 was good correspondence in some cases at the species level: *P. pungens* and *P. multiseries*
381 were sister clades. *P.multistriata* and *P. australis* are sister clades using larger D1-D3 (Lim et
382 al. 2013) but formed one clade in this study. *P. hasleana* and *P. calliantha* are sister taxa both
383 in this study and that of Lim et al. (2013). *P. fraudulenta* and *P. subfraudulenta* are not sister
384 clades but both are monophyletic.
385
386 ML phylogeny of longer reads derived from the CLS dataset of the D1-186F, D1-548R LSU
387 fragment (381bp after trimming) confirmed the presence of *P. fraudulenta* and *P. multiseries*
388 in these samples (supplementary Fig.A1). This tree was more robust with *P. multiseries, P.*
389 *pungens* as sister species adjacent to clades containing *P. brasiliana, P. americana, P.*
390 *multistriata, P. seriata, P. australis* that could be separated into their respective species, as
391 also recovered by (Lim et al. 2013) using the longer D1-D3 LSU region. *P. subfraudulenta* is
392 a separate subclade of *P. fraudulenta*, normally these are sister taxa.
393
394 3.2 Environmental species distribution: genetic and SEM identification
395 ML Phylogeny of environmental sequences (Fig. 2, Table 3) of the 320bp trimmed LSU
396 fragment (D1-186F, D1-548R, McDonald et al. (2007) from 11 CPR samples generated a
397 total of 424 sequences; 163 CLS (many identical) and the 261 OTUs from HTS dataset . All
398 sequences are identified in supplementary Table A2). CLS reads from 6 CPR samples were
399 identified either as *P. multiseries*, *P. fraudulenta* or *P. pungens* (Table 4, supplementary Fig.
400 A1). Thirteen groups of OTUs from CLS and HTS dataset were found by ML phylogenetic
401 analysis (Fig. 2, Table). *Pseudo-nitzschia fraudulenta* clade contained 242 environmental
402 sequences (Fig. 2B) and 142 environmental sequences were identified as *P. multiseries* (Fig.
403 2C) both from HTS and CLS. As *P. multiseries* was an usual finding we confirmed its
404 presence in the Sp08C sample by sequencing the same LSU PCR product using a different

405 primer (see materials and methods, Genbank accession-awaiting). A minority of OTUs were
406 related to other species (Table 3): *P. abrensis* and *P. batesiana*, *P. kodamae* and *P. hasleana*,
407 *P. delicatissima* and *P. arenysensis* (2), *P.seriata, P. pungens* (also identified by CLS), *P.*
408 *subfraudulenta, P. galaxiae* (group I, identified by McDonald) and *P. sabit* and *P. galaxiae*
409 (groups II, III, IV, identified by McDonald) and *P. sabit*. Four OTUs could not be identified:
410 Environmental taxa 1 (OTUs 124 132VJ17,190 132VJ17), with 99% BLAST identity to *P.*
411 *hasleana*. OTU 17 83VJ5 (Environmental taxa 2) showed 99% identity by BLAST to *P.*
412 *fraudulenta* that clustered with *P. galaxiae* and *P. delicatissima* identified by (Ruggiero et al.
413 2015). Finally OTU 166 132VJ1, Environmental taxa 3, (98% identify to *P. multiseries*) was
414 sister to several species of *Pseudo-nitzschia* including *P. cuspidata, P. fukuyoi* and *P.*
415 *pseudodelicatissima*, showing 1% similarity by pairwise distance equally to *P. delicatissima,*
416 *P. cf. delicatissima, P. lineola, P. galaxiae, P.multistriata* and *P. pseudodelicatissima*.
417 BLAST identities inaccurate and were not in agreement with phylogeny. HTS generated
418 more diversity than CLS and all species identified using CLS were also generated by HTS
419 (see Table 4) in samples where both methods were used, showing consistency in detection.
420 Even in the two cases where duplicate samples were analysed instead of the same samples, *P.*
421 *fraudulenta* was identified in both samples by CLS and HTS. The only inconsistencies were
422 the detection of a single *P. multiseries* sequence in Au02C(2) by CLS but not in its duplicate
423 sample Au02C. Only CLS identified *P. pungens* in Au08E and *P. fraudulenta* in Sp05C,
424 absent in HTS analysed samples.
425
426 SEM identification was applied to a subset of the 2002-2008 samples used for genetic
427 analysis (Table 4). This revealed typical coastal and open water species composition also
428 found in earlier studies of this region (Table 2). Since no *Pseudo-nitzschia* cells were found
429 in two Eastern samples, additional SEM analysis of samples from 2014 (Table 5, Fig. 3) were
430 carried out from the same area to determine the extent that SEM can uncover species
431 diversity from CPR samples. Both 2002-2008 (Table 4) and 2014 (Table 5) samples showed
432 typical coastal and open water species compositions compared to earlier Pacific studies
433 (Table 2) confirming that CPR sampling is representative for *Pseudo-nitzschia.* Central
434 samples from 2002-2008 could be compared with those of 2014 and revealed different
435 communities in which only *P. turgidula* was common to both sets. *P. heimii* and unidentified
436 species were the only taxa identified from 2002-2008 coastal samples and was not present in
437 2014 coastal samples. SEM-identification showed little correspondence with genetic results
438 (Table 4). *P. fraudulenta, P. seriata, P. multiseries* and *P. pungens* were observable by both

SEM and genetics but only one sample (Au08C) showed correspondence by genetics and SEM and only for *P. fraudulenta*. Little seasonal variation was observed from both sets of SEM results (Tables 4 and 5), in contrast to genetic results. *P. turgidula* and *P. inflatula* were found to be exclusively open water species in previous studies but were not found genetically in any sample. Particularly striking was that only one species, *Pseudo-nitzschia turgidula*, was found by SEM in Sp08C sample yet genetic results showed this sample was the most diverse with 10 different genetic taxa. *Pseudo-nitzschia multiseries* was not previously observed in central samples and *Pseudo-nitzschia galaxiae or P. sabit* has not been reported at all for both regions.

3.3 Ocean conditions

The last "full" PDO cycle consisted of a cool phase from 1947 through 1976 followed by a warm phase from 1977 through (at least) the mid-1990s (Mantua et al. 1997, Zhang et al. 1997). In late 1998, the PDO entered a cold phase that lasted 4 years, followed by a warm phase that lasted 3 years (2002 through 2005), neutral until August 2007, and then a 6-year cold phase through 2013 (interrupted briefly by the moderate El Niño in fall/winter of 2009/10). Monthly values of the PDO index from 2000 through 2010 are shown in Fig. 4 sPDO values were weakly positive during Autumn 2002 (Au02) and Spring 2005 (Sp05), and strongly negative during Spring 2008 (Sp08), Summer 2008 (Su08), and Autumn 2008 (Au08; Table 1). Note that even though the sPDO value was weakly positive during Au02, the PDO had just reversed polarity from cool to warm phase and conditions may have been more representative of transitional periods.

Temporal patterns of monthly SST anomalies for grid cells that encompassed locations where CPR samples were collected closely followed the PDO index in both the Central and Eastern regions (Fig. 4; supplementary Fig. A4). No strong differences in local SST variability was apparent between the two regions, except that the cool PDO phase from late 1998 through 2001 was less pronounced in the Eastern region compared to the Central region. Within a region (Central or Eastern), temporal patterns of local SST variability for grid cells that encompassed locations where CPR samples were collected were very similar to one another and responded similarly to warm and cool phases of the PDO (Fig. 4; supplementary Fig. A4). Synoptic snapshots of SST in the NE Pacific Ocean during months when CPR samples were collected are shown in Fig. 5. These plots show the spatial patterns in the monthly average SST values across the regions during the cool (2002

and 2008) and warm (2005) PDO years and for months in the Spring, Summer and Autumn. The synoptic snapshots of the regions during May 2005 and May 2008 allow a direct comparison of a warm and cool PDO year, respectively, with the average SST across both regions ~2.9°C cooler in 2008 (Fig. 5C, D). During all months, the Central region was always cooler than the Eastern region, and Southern waters were generally warmer compared to Northern waters within the study area (Fig. 5). A strong seasonal pattern is also evident whereby SST is cooler in the spring compared to the summer and autumn (Fig. 5).

3.4 CPR diatom community analysis

Comparing diversity of HTS-generated OTUs between samples (Fig. 6) revealed Eastern samples dominated by *P. fraudulenta* whilst Central samples were more variable. *P. fraudulenta* diversity was present in eight of the nine HTS samples and was common in all coastal (Eastern Pacific) samples, except for Au08E. A large proportion of *P. fraudulenta* OTUs was observed in Au02C. By contrast, *P. multiseries* OTU diversity was generally dominant when *P. fraudulenta* was rare. *P. multiseries* was common in Spring and Autumn samples. Three samples contained a large proportion of *P. multiseries* OTUs (Sp05C, Au08C, Au08E). A small proportion of *P. multiseries* OTUs were present in Au02C, Sp08C, Sp05E, and Su08E. Endemic diversity was observed within *P. fraudulenta* and *P. multiseries* (Fig. 2B and 2C, respectively). Six *P. fraudulenta* environmental OTU clades were found from single samples, from Au02C (2 clades), or Sp05NE (2 clades), Su08E (1 clade) and Au08E (1 clade) whilst one clade contained OTUs from Sp05NE and Sp05E. By contrast, 12 clades of *P. multiseries* environmental OTUs belonged to Au08E (4 clades), Sp05C (4 clades), Au08C (2 clades), and one clade each to Sp08C, Su08E. Five of the *P. multiseries* clades also corresponded with public sequences of strains (KC710107, EF521880, AF417655, KC017458,). These public sequences were related to each other, but globally distributed (Thessen et al. 2009, Ajani et al. 2013). For example there were 31 site differences between Sp05C specific OTU 8 and 183 from 83VJ41 and a clade containing KC710107 and three CLS from 409239201 2 146VJ5 (A6, B5 and B12) from Au08E . No seasonality was detected by SEM in the 2002-2008 or 2014 samples (Tables 4 and 5) but geographical differences were detected between central and eastern regions. Community composition by SEM analysis was very homogenous within each region. Within this small sample set, no clear trend was observed between genetically detected species or population patterns and PDO phase, in which the patterns were more

biogeographical. However, taxa composition in Sp08C stood out as unusually diverse compared to all other samples (Fig. 6). Furthermore, the dominance of *P. multiseries* in A08E, the was different to genetic community composition of other eastern samples. It is worthy to note that no pattern emerged between sample age and species richness that might indicate degradation related alterations, nor were any patterns related to genomic DNA concentration.

Fig. 7 compares the seasonal mean (mean per region over a season for a given year) abundance of total diatoms versus the larger-sized *P. seriata*-sized cells (>3 µm width), and smaller sized *P. delicatissima* sized cells (<3 µm width). No correspondence was found between the seasonal means (Fig. 7) or average cell counts *Pseudo-nitzschia* spp. (data not shown) in samples used for genetic analysis and the number of LSU sequences (Table 4). The abundance of *Pseudo-nitzschia* was not related to the genetic diversity of species found in samples or to SEM detection. The Sp05E sample contained the highest number of *Pseudo-nitzschia* (~7000 cells), mostly consisting of *P. seriata*-sized cells. In comparison, other samples contained fewer than 2000 *Pseudo-nitzschia* cells. In total, *P. seriata*-sized cells were present in 9 out of 16 seasonal means. Four seasonal means recorded *P. delicatissima*-sized cells with only one of those not recording *P. seriata*-sized cells. More *Pseudo-nitzschia* were recorded in the warmer year of 2005 compared to 2002 and 2008. *Pseudo-nitzschia* spp. ranged from 1.5-33% (seasonal mean) of the total diatoms. Total diatoms were generally more abundant in eastern versus central regions, except for Autumn 2005 and 2008, but no geographic pattern was discernible for *Pseudo-nitzschia* seasonal means.

4. Discussion

High-throughput genetic analysis is becoming cheaper and can complement microscopic counts to delineate *Pseudo-nitzschia* in more detail. Our study reveals that HTS sequencing can be utilised on formalin-preserved samples and that these genetic studies are an important addition to microscopic diversity studies in the Pacific, uncovering novel diversity of species and their distributions. Six taxa found using genetics were not previously reported from this region, including three novel genotypes that could not be attributed to current species. Species diversity identified from SEM in these samples were different to those generated by genetics, but similar in composition to previous studies in Table 2, mostly based on microscopic identification. *Pseudo-nitzschia multiseries* was the second most dominant species group found in this study and an unexpected finding as it has not been reported in

541 open Pacific waters. The finding of potentially harmful species in open Pacific waters has
542 implications for monitoring harmful species in Pacific waters and modelling their
543 distribution.
544
545 Both genetic and SEM diversity revealed contrasting species communities from Coastal
546 waters in this region, which are generally iron-rich, nitrate-poor with high phytoplankton
547 productivity compared to open waters communities characterised by lower (and smaller)
548 phytoplankton productivity regions because these waters (called HNLC regions) are iron-
549 poor, but nitrate-rich (Harrison et al. 1999, Ribalet 2010). Studies in NE Pacific waters
550 revealed phytoplankton and *Pseudo-nitzschia* spp. communities were structured by a nutrient
551 gradient from coastal transitional to open water zones, revealing different communities in
552 coastal, transitional and open water zones (Ribalet 2010). The sampling sites from Ribalet et
553 al. (2010) were near to our CPR sampling stations where we found extraordinary intra-
554 species diversity in *Pseudo-nitzschia fraudulenta* and *P. multiseries* by HTS, in which
555 OTUs were exclusively found in only one sample in many cases. This leads us to
556 hypothesize that this may be a species complex with local isolated populations adapted to
557 different regions. On the other hand examples of variants of a globally-distributed *P.*
558 *multiseries* population, described by Ajani et al. (2013), Thessen et al. (2009), was evident,
559 showing local possible local adaptation in a cosmopolitan population. This study confirms
560 physiological findings of (Thessen et al. 2009) revealing *Pseudo-nitzschia* can adapt to
561 multiple environments due to its high genetic variability in which multiple ecotypes of one
562 species succeed each other. DA producing strains of *P. multiseries* (Pn-1) and *P. fraudulenta*
563 (Pn-9, Pn-12) studied by (Thessen et al. 2009) were identical to environmental sequences
564 uncovered in this study. These strains where showed physiological differences to nutrients
565 and interestingly Pn-9 and Pn-12 which were identical, showed different Domoic acid
566 production patterns with growth. Such studies might indicate epigenetic control mechanisms
567 at play and show that defining ecological niches for *Pseudo-nitzschia* requires genetic and
568 physiology studies.
569
570 *Pseudo-nitzschia* abundances determined from microscopic counts of CPR samples were
571 found to be greater in 2005, when SST was warmest, compared to 2002 and 2008 (with
572 cooler SST), indicating greater growth of potentially toxic species with warmer waters
573 potentially a link with PDO. The small sample size prohibits the identification of any
574 conclusive patterns with the PDO or seasonality but our results suggest temperature may

influence species composition. Both central and eastern Autumn samples in 2002 and 2008 were similar in SST and were similar in species composition, despite different nutrient regimes in these regions. Spring samples were the most diverse and harboured all novel diversity particularly Sp08C with the lowest SST. *P. multiseries* appears to prefer cooler waters, its diversity was highest during cooler SST in 2008 and in cooler waters of central regions during the warmer PDO phase in 2005. The transitional state of the waters in Autumn 2002 may have brought about similar habitats that allowed *Pseudo-nitzschia fraudulenta* to thrive in both central and eastern environments, as both regions are connected through by local and long-ranging currents (Harrison et al. 2004, Whitney and Roberts, 2002).

P. multiseries is a large-sized cosmopolitan species (Hasle 2002) that has been reported from coastal locations (Forbes & Denman 1991, Horner & Postel 1993, Hasle 2002, Trainer et al. 2012). Our finding that *P. multiseries* dominated the genetically analysed *Pseudo-nitzschia* species assemblages in two of the four Central region samples is therefore unusual and was not supported with SEM results from partially destroyed samples. *P. multiseries* was however, observed in three eastern samples from 2014, indicating it is captured by the CPR. One possible explanation of its presence is from *P. multiseries* environmental DNA (e-DNA) disseminated from coastal regions but undetectable by microscopy. However, studies have shown e-DNA has a rapid degradation rate in seawater, even small fragments of 100bp can only last days (Thomsen & Willerslev 2015). Thus it would be difficult to conceive e-DNA surviving the approximately 800 kilometres (430 nautical miles) from coastal to open water communities. The possibility of sample contamination remains from DNA of broken cells taken from Eastern samples leaking through on the CPR sample roll or from the formalin tank to contaminate Central samples. CPR samples are collected on a role of silk with another layer of silk sandwiched over the plankton layer. Several layers of silk separate samples collected from open and coastal regions (Richardson et al. 2006). Central and Eastern samples are separated by approximately 430 nautical miles. The longest CPR tow route is 500 nautical miles, requiring 5 metres of silk so these samples are farthest away from each other, separated by no less than 4m of silk on a roll (Richardson et al. 2006). The possibility remains but is remote and minor. The diversity of OTUs in four independent central samples, should be equivalent or more in their eastern counterparts if contamination from the latter was the source of *P. multiseries*. However, this is not the case and furthermore OTUs specific only to Central samples are found not present in Eastern samples from the same tow, making contamination an unlikely option. The alternative explanation is that local populations of *P.*

609 *multiseries* are supported by mesoscale-level Haida eddies currents containing large volumes
610 of water and nutrients transported up to 1000 kilometers from their point of origin to HNLC
611 regions of NE Pacific (Whitney & Robert 2002). These currents could also bring and support
612 local coastal and cosmopolitan populations for extended periods that may create new hybrid
613 forms thus resulting in a mixture of localised and global populations.
614
615 A good match was found between HTS and CLS genetic approaches in terms of species
616 detected. However, the diversity within species from CLS was severely depleted compared to
617 HTS and thus this approach is not recommended for diversity studies. Several potential
618 reasons could cause the lack of congruity between genetic and SEM results. Sampling and
619 processing differences using the two method is likely a main contributor. Diatoms are one of
620 a few examples where genetics and SEM correspond with adequately resolved genetic
621 markers (Malviya 2016). However, SEM is not a high-throughput method that may mean
622 diversity is lost, especially in this case where part of the sample was destroyed for genetic
623 analysis. A second main issue is the lack of resolution of this marker, combined with
624 population structure within several species common to Pacific Eastern waters, such as *P.*
625 *pseudodelicatissima* groups, and conspecific species groups *P. hasleana/P. kodamae, P.*
626 *galaxiae/P.sabit* and *P. delicatissima*/*P. arenysensis* that were resolved with longer LSU
627 region as shown in previous studies (Lim et al. 2012, Lundholm 2012). The use of new
628 technology such as MinION sequencing (Oxford Nanopores Ltd) that directly sequences
629 long-reads of genomic DNA without an amplification step would reduce bias brought about
630 by PCR analysis of mixed templates (Suzuki 1996, Kalle et al. 2014) and allow better
631 delineation of species with longer reads. The lack of public reference sequences especially for
632 Pacific open water species such as *P. turgidula* and *P. inflatula* could be an additional factor
633 in identification. Increased database representation would improve the phylogeny delineating
634 species or populations within species improving the identification of unknown environmental
635 sequences. For example, *P. turgidula* public sequence appears outside the core group
636 identified, which may be a misidentification or a genuine population variant of that species.
637 Traditional pairwise alignment sequence dissimilarity (PSD, as it is referenced in (Nguyen
638 2016) clustering, similar to the method used in this paper, has been shown to create poor
639 OTU clusters. Better clustering methods-using curated and representative sequence databases
640 to identify OTUs within the bioinformatic pipeline would also improve OTU retrieval
641 process.
642

643 It is clear from this study that better characterisation and more comparative work with both
644 genetics and SEM would benefit characterisation of *Pseudo-nitzschia* in this region.
645 Nevertheless new species to this region have been uncovered using HTS approach on
646 archival formalin-preserved samples. *P. galaxiae, P. sabit,* and *P. kodamae* or
647 *P.subfraudulenta* have not been described in Northern Pacific Eastern waters whilst *P.*
648 *hasleana* has only been identified in coastal waters of this region (Table 2) that demonstrates
649 that diversity in NE Pacific is under-characterised. All but *P. hasleana* (Lundholm 2012)
650 have mainly been identified in warmer waters (Lundholm 2002, Teng et al. 2014, Teng
651 2015). Our findings show a broader distribution range of *Pseudo-nitzschia* species in Pacific
652 waters. Open water species deserved further study to capture and culture representatives to
653 determine their environmental preferences. Their response to nutrients and temperature make
654 them valuable indicators of ocean health.
655

656 **5. Acknowledgements**

657 We wish to thank the officers and crew of the M/V Skaubryn operated by Seaboard
658 International for towing the CPR from their ship and to SAHFOS staff for providing data for
659 this paper. This project was funded by the SAHFOS associated researcher scheme. Sample
660 collection was funded by the North Pacific CPR Consortium managed by the North Pacific
661 Marine Science Organisation (PICES). Funding for sample collection was provided by the
662 North Pacific Research Board, Canadian Department of Fisheries and Oceans and the Exxon
663 Valdez Oil Spill Trustee Council. We thank the reviewers of this manuscript for their helpful
664 advice.
665

666 **References**

667 Ajani P, Murray S, Hallegraeff G, Lundholm N, Gillings M, Brett S, Armand L (2013) The
668 diatom genus Pseudo-nitzschia (Bacillariophyceae) in New South Wales, Australia:
669 morphotaxonomy, molecular phylogeny, toxicity, and distribution. Journal of
670 Phycology 49:765-785
671 Al-Kandari M (2012) Molecular characterisation of diversity in an exceptional harmful algal
672 bloom forming species, *Karenia mikimotoi*, in the Celtic Sea shelf break region. PhD,
673 University of Plymouth, Plymouth
674 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search
675 tool. Journal of Molecular Biology 215:403-410
676 Amato A, Kooistra WH, Ghiron JH, Mann DG, Proschold T, Montresor M (2007)
677 Reproductive isolation among sympatric cryptic species in marine diatoms. Protist
678 158:193-207
679 Anderson DM (1995) The ecology and oceanography of harmful algal blooms. A national
680 research agenda. Woods Hole Oceanographic Institution, Woods Hole, MA

681 Andree KB, Fernández-Tejedor M, Elandaloussi LM, Quijano-Scheggia S, Sampedro N,
682 Garcés E, Camp J, Diogène J (2011) Quantitative PCR Coupled with Melt Curve
683 Analysis for Detection of Selected Pseudo-nitzschia spp. (Bacillariophyceae) from the
684 Northwestern Mediterranean Sea. Appl Environ Microbiol 77:1651-1659
685 Auro ME (2007) Nitrogen dynamics and toxicity of the pennate diatom *Pseudonitzschia*
686 *cuspidata*: a field and laboratory study. Masters Thesis, San Francisco State
687 University, San Francisco, CA
688 Batten SD, Clark. R., Flinkman J, Hays G, John E, John AWG, Jonas T, Lindley JA, Stevens
689 DP, Walne A (2003a) CPR sampling: the technical background, materials and
690 methods, consistency and comparability. Progress in Oceanography 58:193-215
691 Batten SD, Hyrenbach, K.D., Sydeman, W.J., Morgan, K.H., Henry, M.F., Yen, P.Y., Welch,
692 D.W. (2006) Characterising Meso-Marine Ecosystems of the North Pacific. . Deep
693 Sea Research II 53:270-290
694 Batten SD, Walne AW, Edwards M, Groom SB (2003b) Phytoplankton biomass from
695 continuous plankton recorder data: an assessment of the phytoplankton colour index. J
696 Plankton Res 25:697-702
697 Botsford LW, Castilla JC, Peterson CH (1997) The Management of Fisheries and Marine
698 Ecosystems. Science 277 277
699 Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N,
700 Gonzalez Pena A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D,
701 Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J,
702 Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J., R.
703 K (2010) QIIME allows analysis of high-throughput community sequencing data.
704 Nature Methods
705 Dyson K, Huppert DD (2010) Regional economic impacts of razor clam beach closures due
706 to harmful algal blooms (HABs) on the Pacific coast of Washington. . Harmful Algae
707 9: =264–271
708 Fernandes LF, Hubbard KA, Richlen ML, Smith J, Bates SS, Ehrman J, Léger C, Mafra LL,
709 Kulis D, Quilliam M, Libera K, McCauley L, Anderson DM (2014) Diversity and
710 toxicity of the diatom Pseudo-nitzschia Peragallo in the Gulf of Maine, Northwestern
711 Atlantic Ocean. Deep Sea Research Part II: Topical Studies in Oceanography
712 103:139-162
713 Fernandes LF, Hubbard, K.A., Richlen, M.L., Smith, J., Bates, S. S., Ehrman, J. et al. (2014)
714 Diversity and toxicity of the diatom *Pseudo-nitzschia* Peragallo in the Gulf of Maine,
715 Northwestern Atlantic Ocean. . Deep-sea research Part II, Topical studies in
716 oceanography 103:139-162
717 Field D, Tiwari B, Booth T, Houten S, Swan D, Bertrand N, Thurston M (2006) Open
718 Software for biologists: from famine to feast. . Nature Biotechnology 24:801 - 803
719 Forbes JR, Denman KL (1991) Distribution of *Nitzschia pungens* in coastal waters of British
720 Columbia. Can J Fish Aquat Sci 48:960–967
721 Fryxell GA, Célia-Villac M, Shapiro LP (1997) The occurrence of the toxic diatom genus
722 *Pseudo-nitzschia* (Bacillariophyceae) on the West Coast of the USA, 1920-1996:
723 a review. . Phycologia 36:419-437
724 Garćia-Mendoza E, Rivas D, Olivos-Ortiz A, Almazán-Becerril A, Castañeda-Vega C (2009)
725 A toxic *Pseudo-nitzschia* bloom in Todos Santos Bay, northwestern Baja California,
726 Mexico. Harmful Algae 8:493-503
727 Gómez F, Claustre H, Raimbault P, Souissi S (2007) Two High-Nutrient Low-Chlorophyll
728 phytoplankton assemblages: the tropical central Pacific and the offshore Peru-Chile
729 Current. . Biogeosciences 4:1101-1113

730 Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis
731 program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98
732 Hare SR, Mantua, N.J., Francis, R.C. (1999) Inverse Production Regimes: Alaska and West
733 Coast Pacific Salmon. . Fisheries 24:6-14
734 Harrison PJ, Boyd PW, Varela DE, Takeda S, Shiomoto A, Odate T (1999) Comparison of
735 factors controlling phytoplankton productivity in the NE and NW subarctic Pacific
736 gyres. Progress in Oceanography 43:205-234
737 Hasle GR (1993) Nomenclatural notes on marine planktonic diatoms.The family
738 Bacillariaceae. Nova Hedwigia, Beiheft 106:315-321
739 Hasle GR (2002) Are most of the domoic acid-producing species of the diatom genus
740 *Pseudo-nitzschia* cosmopolites? Harmful Algae 1:137-146
741 Hernández-Becerril DU (1998) Species of the planktonic diatom genus *Pseudo-nitzschia* of
742 the Pacific coasts of Mexico. Hydobiologica 379:77-84
743 Hernández-Becerril DU, Bravo-Sierra, E., Aké-Castillo, J.A. (2007) Phytoplankton on the
744 western coasts of Baja California in two different seasons in 1998. Sci Mar 71:735-
745 743
746 Horner RA, Postel JR (1993) Toxic diatoms in western Washington waters (U.S. West
747 Coast). Hydrobiol 269/270:197-205
748 Hubbard KA, Rocap GE, Armbrust V (2008) Inter and intraspecific community structure
749 within the diatom genus *Pseudo-nitzschia* (Bacillariophyceae). J Phycol 44
750 JISAO (2014) University of Washington Pacific Decadal Oscillation (PDO) Accessed 10 July
751 2014. http://jisao.washington.edu/pdo/PDO.latest
752 Karmakar S, Harcourt EM, Hewings DS, Scherer F, Lovejoy AF, Kurtz DM,
753 Ehrenschwender T, Barandun LJ, Roost C, Alizadeh AA, Kool ET (2015)
754 Organocatalytic removal of formaldehyde adducts from RNA and DNA bases. Nat
755 Chem 7:752-758
756 Letunic I, Bork, P. (2016) Interactive Tree Of Life (iTOL) v3: an online tool for the display
757 and annotation of phylogenetic and other trees. Nucleic Acids Res
758 Lim H-C, Leaw C-P, Su SN-P, Teng S-T, Usup G, Mohammad-Noor N, Lundholm N, Kotaki
759 Y, Lim P-T (2012) Morphology and molecular characterization of *Pseudo-nitzschia*
760 (Bacillariophyceae) from Malaysian Borneo, including the new species *Pseudo-*
761 *nitzschia circumpora* sp. nov. Journal of Phycology 48:1232-1247
762 Lim HC, Teng ST, Leaw CP, Lim PT (2013) Three novel species in the *Pseudo-nitzschia*
763 *pseudodelicatissima* complex: *P. batesiana* sp. nov., *P. lundholmiae* sp. nov., and *P.*
764 *fukuyoi* sp. nov. (Bacillariophyceae) from the Strait of Malacca, Malaysia. . J Phycol
765 49:902–916.
766 Lundholm N (2012) Cryptic and pseudo-cryptic diversity in diatoms-with descriptions of
767 *Pseudo-nitzschia hasleana* sp. nov and *P. fryxelliana* sp. nov. J Phycol 48:436-454
768 Lundholm N, Daujberg N, Moestrup Ø (2002) Phylogeny of the Bacillariaceae with emphasis
769 on the genus *Pseudo-nitzschia* (Bacillariophyceae) based on partial LSU rDNA. Eur J
770 Phycol 37:115-134.
771 Lundholm N, Moestrup Ø, Hasle GR, Hoef-Emden K, (2003) A study of the *Pseudo-*
772 *nitzschia pseudodelicatissima/cuspidata* complex (Bacillariophyceae): what is *P.*
773 *pseudodelicatissima*? J Phycol 39:797–813.
774 Lundholm N, Moestrup Ø, Kotaki Y, Hoef-Emden K, Scholin C, Miller P (2006) Inter- and
775 Intraspecific Variation of the *Pseudo-nitzchia delicatssima* Ccomplex
776 (Bacillariophyceae) illustrated by rRNA probes, morphological data and phylogenetic
777 analysis J Phycol 42:464-481.

778 Lundholm N, Moestrup, Ø (2002) The marine diatom *Pseudo-nitzschia galaxiae* sp. nov.
779 (Bacillariophyceae): morphology and phylogenetic relationships. Phycologia 41 594-
780 605
781 Malviya S, Scalco, E., Audic, S., Vincent, F., Veluchamy, A. et al. (2016) Insights into global
782 diatom distribution and diversity in the world's ocean. PNAS Early edition:1-10
783 Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997) A Pacific Interdecadal
784 Climate Oscillation with impacts on salmon production. Bulletin of the American
785 Meteorological Society 78:1069-1079
786 Marchetti A, Maldano MT, Lane ES, Harrison PJ (2006) Iron requirements of the pennate
787 diatom *Pseudo-nitzschia*: Comparison of oceanic (high-nitrate, low-chlorophyll
788 waters) and coastal species. Limnol Oceanog 51:2092-2101
789 McCabe RM, Hickey BM, Kudela RM, Lefebvre KA, Adams NG, Bill BD, Gulland FMD,
790 Thomson RE, Cochlan WP, Trainer VL (2016) An unprecedented coastwide toxic
791 algal bloom linked to anomalous ocean conditions. Geophys Res Lett
792 McDonald SM, Sarno D, Zingone A (2007) Identifying *Pseudo-nitzschia* species in natural
793 samples using genus-specific PCR primers and clone libraries. Harmful Algae 9:849-
794 860
795 McQuatters-Gollop A, Edwards M, Helaouët P, Johns DG, Owens NJP, Raitsos DE,
796 Schroeder D, Skinner J, Stern RF (2015) The Continuous Plankton Recorder survey:
797 how can long-term phytoplankton datasets deliver Good Environmental Status? Estua
798 Coast Shelf S 162:88-97
799 Nguyen N-P, Warnow, T., Pop, M., White, B. (2016) A perspective on 16S rRNA operational
800 taxonomic unit clustering using sequence similarity. npj Biofilms and Microbiomes
801 2:16004
802 Orsini L, Procaccini G, Sarno D, Montresor M (2004) Multiple rDNA ITS-types within the
803 diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae) and their relative
804 abundances across a spring bloom in the Gulf of Naples. . Mar Ecol Prog Ser 271:87-
805 98
806 Orsini L, Sarno, D., Procaccini, G., Poletti, R., Dahlman, J., Montresor, M. (2002) Toxic
807 *Pseudo-nitzschia multistriata* (Bacillariophyceae) from the Gulf of Naples:
808 morphology, toxin analysis and phylogenetic relationships with other Pseudo-
809 nitzschia species. Eur J Phycol 37:247-257
810 Paireder S, Werner B, Bailer J, Werther W, Schmid E, Patzak B, Cichna-Markl M (2013)
811 Comparison of protocols for DNA extraction from long-term preserved formalin fixed
812 tissues. Anal Biochem 439 152–160
813 Penna A, Casabianca S, Perini F, Bastianini M, Riccardi E, Pigozzi S, Scardi M (2012) Toxic
814 Pseudo-nitzschia spp. in the northwestern Adriatic Sea: characterization of species
815 composition by genetic and molecular quantitative analyses. Journal of Plankton
816 Research
817 Percopo I, Ruggiero MV, Balzano S, Gourvil P, Lundholm N, Siano R, Tammilehto A,
818 Vaulot D, Sarno D, Mock T (2016) Pseudo-nitzschia arctica sp. nov., a new
819 cold-water cryptic Pseudo-nitzschia species within the P. pseudodelicatissima
820 complex. Journal of Phycology 52:184-199
821 Perl TM, Teitelbaum J, Hockin J, Todd EC (1990) Domoic acid toxicity. Panel discussion:
822 definition of the syndrome. Canada Diseases Weekly Report 16 (Suppl 1E):41-45
823 Raitsos DE, Reid, P.C., Lavender, S.J., Edwards, M., Richardson, A.J. (2005) Extending the
824 SeaWIFS chlorophyll data set back 50 years in the northeast Atlantic. Geophys Res
825 Lett 32:LO6603
826 Ribalet F, Marchetti, A., Hubbard, K. A., Brown, K., Durkin, C. A., Morales, R., Robert, M.,
827 Swallwell, J.E., Tortell, P.D., Armbrust, E. V. (2010) Unveiling a phytoplankton

828 hotspot at a narrow boundary between coastal and offshore waters. . PNAS
829 107:16571-16576
830 Richardson AJ, Walne AW, John AWG, Jonas T.D., Lindley J.A., Sims DW, Stevens D, Witt
831 M (2006) Using continuous plankton recorder data. Prog Oceanogr 68:27-74
832 Ripley SJ, Baker AC, Miller PI, Walne AW, Schroeder DC (2008) Development and
833 validation of a molecular technique for the analysis of archived formalin-preserved
834 phytoplankton samples permits retrospective assessment of *Emiliania huxleyi*
835 communities. J Microbiol Methods 73:118-124
836 Robinson GA, Hiby AR (1978) The Continuous Plankton Recorder Survey. In: Sournia A
837 (ed) Phytoplankton Manual. UNESCO., Paris
838 Ruggiero MV, Sarno D, Barra L, Kooistra WHCF, Montresor M, Zingone A (2015) Diversity
839 and temporal pattern of *Pseudo-nitzschia* species (Bacillariophyceae) through the
840 molecular lens. Harmful Algae 42:15-24
841 Scholin CA, Gulland F, Doucette GJ, Benson S, Busman M, Chavez FP, Cordaro J, DeLong
842 R, De Vogelaere A, Harvey J, Haulena M, Lefebvre K, Lipscomb T, Loscutoff S,
843 Lowenstine LJ, Marin Iii R, Miller PE, McLellan WA, Moeller PDR, Powell CL,
844 Rowles T, Silvagni P, Silver M, Spraker T, Trainer V, Van Dolah FM (2000)
845 Mortality of sea lions along the central California coast linked to a toxic diatom
846 bloom. Nature 403:80-84
847 Scholin CA, Herzog, M., Sogin, M., Anderson, D.M. (1994) Identification of Group- and
848 Strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae).
849 II. Sequence analysis of a fragment of the LSU rRNA gene. J Phycol 30:999-1011
850 Scholz MB, Lo. CC, Chain PSG (2012) Next generation sequencing and bioinformatic
851 bottlenecks: the current state of metagenomic data analysis. Curr Opin Biotechnol
852 23:9-15
853 Silver MW, Bargu S, Coale SL, Benitez-Nelson CR, Garcia AC, Roberts KJ, Sekula-Wood
854 E, Bruland KW, Coale KH (2010) Toxic diatoms and domoic acid in natural and iron
855 enriched waters of the oceanic Pacific. . Proc Natl Acad Sci USA 107:20762-20767
856 Stehr CM, Connell, L., Baugh, K. A., Bill, B. D., Adams, N. G., Trainer, V. L. (2002)
857 Morphological, toxicological, and genetic differences among *Pseudo-nitzschia*
858 (Bacillariophyceae) species in inland embayments and outer coastal waters of
859 Washington State, USA. . J Phycol 38:55–65
860 Stonik IV, Orlova, T.Y., Lundholm, N. (2011) Diversity of *Pseudo-nitzschia* H. Peragallo
861 from the western North Pacific. Diatom Research 26:121-134
862 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular
863 Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution
864 30:2725-2729
865 Teng ST, Lim HC, Lim PT, Dao VH, Bates SS, Leaw CP (2014) Pseudo-nitzschia kodamae
866 sp. nov. (Bacillariophyceae), a toxigenic species from the Strait of Malacca, Malaysia.
867 Harmful Algae 34:17-28
868 Teng ST, Lim, P.T., Lim, H.C., Rivera-Vilarelle, M., Quijano-Scheggia, S., *et al.* (2015) A
869 non- toxigenic but morphologically and phylogenetically distinct new species of
870 *Pseudo-nitszschia*, *P. sabit* sp. nov. (Bacillariophyceae. J Phycol 51:706-725
871 Thessen AE, Bowers HA, Stoecker DK (2009) Intra- and interspecies differences in growth
872 and toxicity of *Pseudo-nitzschia* while using different nitrogen sources. Harmful
873 Algae 8:792-810
874 Thomsen PF, Willerslev E (2015) Environmental DNA – An emerging tool in conservation
875 for monitoring past and present biodiversity. Biol Cons 183:4-18

Trainer VL, Bates SS, Lundholm N, Thessen AE, Cochlan WP, Adams NG, Trick CG (2012) *Pseudo-nitzschia* physiological ecology, phylogeny, toxicity, monitoring and impacts on ecosystem health. Harmful Algae 14:271-300

Trainer VL, Hickey B. M., Horner RA (2002) Biological and physical dynamics of domoic acid production off the Washington U.S.A. coast. . Limnol Oceanogr 47:1438-1446

Trainer VL, Suddleson M (2005) Monitoring approaches for early warning of domoic acid events in Washington State Oceanography 18:228–237

Trick CG, Bill BD, Cochlan WP, Wells ML, Trainer VL, Pickell LD (2010) Iron enrichment stimulates toxic diatom production in high-nitrate, low-chlorophyll areas. . Proc Natl Acad Sci USA 107:5887-5892

Vezzulli L, Brettar I, Pezzati E, Reid PC, Colwell RR, Hofle MG, Pruzzo C (2012) Long-term effects of ocean warming on the prokaryotic community: evidence from the *Vibrio*s. The ISME Journal 6:21-30

Vezzulli L, Grande C, Reid PC, Helaouet P, Edwards M, Hofle MG, Brettar I, Colwell RR, Pruzzo C (2016) Climate influence on *Vibrio* and associated human diseases during the past half-century in the coastal North Atlantic. PNAS

White TJ, Bruns, T., Lee, S., Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand, D.H., Sninsky, J.J., White, T.J. (ed) PCR Protocols: a guide to methods and applications. Academic Press, New York, U.S.A.

Whitney FA, Robert M (2002) Structure of Haida eddies and their transport of nutrient from coastal margins into the NE Pacific Ocean. . J Oceanogr 58:715-723

Work TM, Barr B, Beale AM, Fritz L, Quilliam MA, Wright JLC (1993) Epidemiology of Domoic Acid Poisoning in Brown Pelicans (*Pelecanus occidentalis*) and Brandt's Cormorants (*Phalacrocorax penicillatus*) in California. J Zoo Wildl Med 24:54-62

Zamudio-Resendiz ME, Gónzalez-Rivas D, Meave del C ME (2014) Evaluation of *Pseudo-nitzschia* spp. in a tropical bay of the Mexican Pacific. In: Kim HG, B. Reguera, G.M. Hallegraeff, C.K. Lee, M.S. Han and J.K. Choi. (ed) 15th International Conference of Harmful Algae, Book ISBN 978-87-990827-4-2. International Society for the Study of Harmful Algae, Changwon, Korea

Zhang Y, Wallace JM, Battisti DS (1997) ENSO-like interdecadal variability: 1900-93. J Climate 10:1004-1020

Figures

Fig. 1. Continuous Plankton Recorder samples collected from in 2002 (+), 2005 (o), 2008 (Δ), and 2014 (×; SEM). Samples were selected from 1-2 transects conducted during different seasons. Grid cells indicate the 1° latitude by 1° longitude spatial resolution of satellite-derived SST used in this study. Gray-shaded grid cells (labelled A-H) contain the ten CPR samples used for molecular analysis represented by shaded symbols ✚, ● and ▲. These correspond to the time series of SST anomalies shown in Supplementary Fig. A4. Details of samples subjected to molecular analysis are listed in Table 1 and Table 4. Central and Eastern regions are bisected by the -134˚E longitude line (-136˚E for the northern transect) for the community composition analyses.

Fig. 2. LSU Maximum Likelihood (ML) phylogeny from a 439bp alignment of partial LSU fragment from public reference sequences and environmental *Pseudo-nitzschia* sequences from 11 CPR samples (A). Some clades have been collapsed for clarity, those marked with an asterix also have environmental sequences. Genetic distances are not shown here for clarity but are shown in supplementary Fig.A3). Expanded subtrees with genetic distances show microdiversity of *Pseudo-nitzschia fraudulenta* (panel B) and *Pseudo-nitzschia multiseries with Pseudo-nitzschia pungens* (panel C) . Grey boxes and indicate clades that correspond to Fig. 2B. Asterix indicates sequences recovered together in Fig, 2C.Environmental sequences are in bold type. Bootstrap values over 70 and branch length are shown by their respective clades. Genetic distances of the whole tree are indicated on the top left corner.

Fig. 3. SEM images of (A) *Pseudo-nitzschia fraudulenta* and (B) *Pseudo-nitzschia inflatula* (C) *Pseudo-nitzschia pungens* in 2014 CPR samples. See Table 4 for locations.

Fig. 4. Time series of monthly values of the PDO index from 2000 through 2010 indicating warm, cool, and neutral phases. Asterisks indicate months when CPR samples used in the molecular analyses were collected; October 2002 (Au02E, Au02C and Au02C(2)); April 2005 (Sp05NE); May 2005 (Sp05E and Sp05C); May 2008 (Sp08C); July 2008 (Su08E); and September 2008 (Au08E and Au08C).

Fig. 5. Spatial variability in monthly averages of satellite-derived SST in the NE Pacific during months when the CPR samples used in the molecular analyses were collected. Maps show contoured SST during (A) October 2002 (Au02E, Au02C and Au02C(2)); (B) April 2005 (Sp05NE); (C) May 2005 (Sp05E and Sp05C); (D) May 2008 (Sp08C); (E) July 2008 (Su08E); and (F) September 2008 (Au08E and Au08C).

Fig. 6. Diversity of Environmental *Pseudo-nitzschia* OTUs diversity per taxa found in samples analysed from HTS environmental reads, clustered at 99% and 90% identity.

Fig. 7. Average cell counts of all diatoms (open bars) compared to *Pseudo-nitzschia seriata*-sized cells and *Pseudo-nitzschia delicatissima*-sized cells in each region. Asterix indicates that there are genetic data available from at least one sample from each seasonal mean.

Tables

Table 1. Summary of CPR samples used in this study and the seasonal mean values of the PDO index (sPDO) for the season in which the sample was collected. Samples are provided with codes to denote the season (Autumn, Summer, Spring), year (2002, 2005, 2008) and region (Eastern or Central) that they come from and are listed following their longitudinal position. Lat= Latitude, Long= Longitude. sPDO values for autumn ("Au") are the mean of September, October, and November; summer ("Su") is the mean of June, July, and August; and spring ("Sp") is the mean of March, April, May. Samples 21VJ5 (Au02E(2)) and 21VJ45 (Au02C(2)) are regional duplicates of 21VJ1 and 21VJ41

CPR Sample	Month	Year	Lat (°N)	Long (°E)	Location	Code	sPDO
21VJ1	10	2002	48.71	-125.42	Eastern	Au02E	0.79
21VJ5	10	2002	48.71	-125.42	Eastern	Au02E(2)	0.79
139VJ1	7	2008	48.76	-125.99	Eastern	Su08E	-1.57
146VJ5	9	2008	48.7	-126.17	Eastern	Au08E	-1.52
83VJ5	5	2005	48.88	-126.48	Eastern	Sp05E	0.69
77VJ7	4	2005	54.97	-134.97	Eastern	Sp05NE	1.42
21VJ41	10	2002	51.75	-134.61	Central	Au02C	0.79
132VJ17	5	2008	48.76	-136.79	Central	Sp08C	-1.20
146VJ37	9	2008	49.92	-134.11	Central	Au08C	-1.52
83VJ41	5	2005	51.3	-135.02	Central	Sp05C	0.69
21VJ45	10	2002	51.95	-135.64	Central	Au02C(2)	0.79

965 **Table 2.** Coastal and open ocean *Pseudo-nitzschia* species reported from the Pacific Ocean in the literature with their approximate dimensions.
966 LM = light microscopy; TEM = transmission electron microscopy; SEM = scanning electron microscopy. Shaded cells indicate species that
967 overlap both coastal and open ocean niches. Question mark indicates uncertain identification in citation.
968

***Pseudo-nitzschia* sp.**	**Niche**	**Pacific Ocean region**	**Width (µm)**	**Length (µm)**	**Identification method**	**Reference**
P. pungens	Coastal	USA (WA, OR, CA); Peru; Mexico; SE Pacific	2.4-5.3	74-174	Genetic; LM; TEM	(Fryxell et al. 1997, Hubbard et al. 2008, Stonik 2011, Trainer et al. 2012)
P. multiseries	Coastal	USA (WA, CA); Peru; SE Pacific	3.4-6.0	68-140	Genetic; LM; TEM	(Fryxell et al. 1997, Hubbard et al. 2008, Stonik 2011, Trainer et al. 2012)
P. seriata	Coastal	USA (WA, CA); Peru; SEPacific	5.5-8.0	75-160	Genetic; LM; TEM	(Gómez et al. 2007, Hubbard et al. 2008, Stonik 2011)
P. australis	Coastal	USA (WA, OR, CA)	6.5-8.0	75-144	Genetic, SEM	(Fryxell et al. 1997, Hubbard et al. 2008, Garćia-Mendoza et al. 2009, Trainer et al. 2012)
P. subpacifica	Coastal	USA (WA, CA)	5-7	33-70	Genetic, LM	(Fryxell et al. 1997, Hubbard et al. 2008)
P. cuspidata	Coastal	USA (CA, WA)	~3	30-80	Genetic; LM	(Fryxell et al. 1997, Auro 2007, Lundholm 2012, Trainer et al. 2012)
P. calliantha	Coastal	USA (WA); Peru; Western Pacific	4-6	30-72	LM; TEM	(Marchetti et al. 2006, Stonik 2011)
P. multistriata	Coastal	Peru; SE Pacific	2.5-3.8	38-65	LM; TEM	(Gómez et al. 2007, Stonik 2011)
P. obtusa	Coastal	Peru; SE Pacific	4.5-5.5	61-100	LM; TEM	(Gómez et al. 2007, Stonik 2011)
P. cf. *caciantha*	Coastal	Peru;SE Pacific	3.5-5	53-75	LM; TEM	(Gómez et al. 2007, Stonik 2011)
P. americana	Coastal	Peru	~3	16-40	LM	(Gómez et al. 2007)

P. subfraudulenta	Coastal	Mexico, USA (CA)	3.7-7.0	65-133	LM	(Fryxell et al. 1997, Zamudio-Resendiz et al. 2014)
P. hasleana	Coastal	USA (WA)	1.5-2.8	37-79	Genetic, SEM	(Lundholm 2012)
P. australis	Coastal	USA (WA, OR, CA)	6.5-8.0	75-144	Genetic, SEM	(Fryxell et al. 1997, Hubbard et al. 2008, Garćia-Mendoza et al. 2009, Trainer et al. 2012)
	Open	NE Pacific	6.5-8.0	75-144	SEM	Trainer et al. 2012
P. fraudulenta	Coastal	USA (WA); Peru; SE Pacific	4.5-10.0	50-119	Genetic; LM; SEM; TEM	(Horner & Postel 1993, Fryxell et al. 1997, Hubbard et al. 2008, Stonik 2011)
	Open	NE Subarctic Pacific (Station AL)	4.5-10.0	50-119	LM, SEM; TEM	(Silver et al. 2010)
P. pseudodelicatissima	Coastal	USA (WA); Mexico	1.3-2.5	59-140	LM; SEM	(Fryxell et al. 1997, Trainer et al. 2002, Zamudio-Resendiz et al. 2014)
	Open	NE Subarctic Pacific (Station AL)	1.3-2.5	59-140	LM; SEM; TEM	(Silver et al. 2010)
P. delicatissima	Coastal	USA (WA); SE Pacific	1-2	40-76	Genetic; LM; TEM	(Fryxell et al. 1997, Hubbard et al. 2008, Stonik 2011, Trainer et al. 2012)
	Open	SE Pacific HNLC	1-2	40-76	LM	(Gómez et al. 2007)
P. heimii/ P. cf. *heimii*	Coastal	USA (WA); Peru; SE Pacific	4-6	67-120	LM; TEM	(Fryxell et al. 1997, Gómez et al. 2007, Stonik 2011)
	Open	NE Pacific (Ocean Station PAPA); NE Subarctic Pacific (Station AL)	4-6	67-120	LM; SEM; TEM	(Marchetti et al. 2006, Silver et al. 2010)
P. lineola	Coastal		1.8-2.7	56-112	Genetic; LM; SEM	(Fryxell et al. 1997, Hernández-Becerril 1998, 2007, Garćia-Mendoza et al.

						2009)
	Open	NE Subarctic Pacific (Station AL)	1.8-2.7	56-112	LM; SEM; TEM	(Silver et al. 2010)
P. turgidula/ *P.* cf. *turgidula*	Coastal	California	1.3-2.5	30-80	LM	(Fryxell et al. 1997)?
	Open	NE Pacific (Ocean Station PAPA); NE Subarctic Pacific (Station AL)	1.3-2.5	30-80	LM; SEM; TEM	(Silver et al. 2010, Trick et al. 2010)
P. grannii, P. cf. *grannii*	Open	NE Pacific (Ocean Station PAPA); NE Subarctic Pacific (Station AL)	1.5-2.5	25-79	LM; SEM; TEM	(Silver et al. 2010, Trick et al. 2010)
P. dolorosa	Open	NE Pacific (Ocean Station PAPA)	2-3.2	30-59	LM; TEM	(Marchetti et al. 2006)
P. inflatula	Coastal	USA (CA)	1.5-2.5	6-100	LM	(Fryxell et al. 1997)?
	Open	NE Subarctic Pacific (Station AL)	1.5-2.5	6-100	LM; SEM; TEM	(Silver et al. 2010)

969

970

971 **Table 3:** Species groups identified using in this study from ML phylogeny (Fig. 2) and related to previous studies showing the number of
972 environmental sequences from this study that corresponds to each group.

MP clade group name	species in clade	Strain correspondence with previous studies	Number environmental sequences
P. abrensis, P. batesiana	*P. abrensis, P. batesiana*	No	2
P. dolorosa, P. micropora	*P. dolorosa, P. micropora, P. cf. delicatissima (2)*	(Lim et al. 2012, Ajani et al. 2013)	0

P. pseudodelicatissima	*P. pseudodelicatissima*	(Orsini 2002)	0
P. mannii	*P. mannii*	(Lim et al. 2012, Lundholm 2012, Ajani et al. 2013)	0
P. kodamae, P. hasleana	*P. kodamae, P. hasleana*	(Lundholm 2012, Ajani et al. 2013)	11
P. delicatissima, P .arenysensis (clade 2)	*P. delicatissima, P. arenysensis, Pseudonitzschia sp.*	*P. delicatissima* not confirmed by other studies. One sequence of *P. arenysensis* was formerly *P. delicatissima* (Ajani et al. 2013), not in same clade as other *P. arenysensis. Pseudo-nitzschia* sp. identified as *Pseudo-nitzschia* new genotype, sister to *P. delicatissima* (McDonald et al. 2007)	5
P. multistriata, P. australis	*P. multistriata, P. australis*	(Lim et al. 2012, Ajani et al. 2013, Lim et al. 2013)	0
P. brasiliana (sensu stricto)	*P. brasiliana*	(Lim et al. 2012, Ajani et al. 2013, Lim et al. 2013)	0
P. linea	*P. linea*	No	0
P. americana	*P. americana*	(Ajani et al. 2013)	0
P. seriata	*P. seriata*	Only one sequence separate from a second sequence but recognised in 3 studies (Lim et al. 2012, Ajani et al. 2013, Lim et al. 2013)	4
P. pungens	*P. pungens, Pseudonitzschia pungens var. aveirensis*	(Lim et al. 2012, Ajani et al. 2013, Lim et al. 2013)	2
P. multiseries	*P. multiseries*	(Lim et al. 2012, Ajani et al. 2013, Lim et al. 2013)	142
P. subfraudulenta	*P. subfraudulenta*	(Lim et al. 2012, Ajani et al.	4

		2013, Lim et al. 2013)	
P. lundholmiae	*P. lundholmiae*	(Lim et al. 2013)	0
P. lineola	*P. lineola*	(Lundholm 2012, Ajani et al. 2013)	0
MVR2015	*Pseudo-nitzschia* sp. MVR2015, Bacillariophyceae MVR2015	No	0
P. inflatula	*P. inflatula*	(Lim et al. 2012, Lundholm 2012)	0
P. delicatissima, P. arenysensis	*P. delicatissima, P. arenysensis, P. pseudodelicatsissima , P. multistriata , P. galaxiae*	*P. delicatissima* and *P. arenysensis* strains confirmed by 4 studies (Orsini 2002, Stehr 2002, Lim et al. 2012, Lundholm 2012). *P. pseudodelicatissima* (Orsini 2002) sister *to P. pseudodelicatissima* group.	0
P. subpacifica, P.heimii	*P. subpacifica, P.heimii*	(Lim et al. 2012, Ajani et al. 2013)	0
P. fraudulenta	*P. fraudulenta*	(Lim et al. 2012, Ajani et al. 2013, Lim et al. 2013)	242
P. fryxelliana	*P. fryxelliana*	(Lim et al. 2012, Lundholm 2012, Ajani et al. 2013, Lim et al. 2013)	0
P. circumpora	*P. circumpora*	(Lim et al. 2012)	0
P. turgidula	*P. turgidula*	(Lundholm 2012, Lim et al. 2013)	0
P. caciantha	*P. caciantha*	No	0
P. arctica	*P. arctica, P. pseudodelicatissima*	*(Percopo et al. 2016)*	0
P. galaxiae I, P. sabit	*P. galaxiae, P. sabit, P. fraudulenta , P. delicatissima*	*P. galaxiae* and *P. sabit* sister clades (Teng 2015). *P. galaxiae* group I strains identified by (McDonald et al.	1

		2007) and *P. galaxiae* identified by (Lundholm 2012). *P. sabit* confirmed by (Teng 2015)	
P. galaxiae II, III, IV, *P. sabit*	*P. galaxiae* II, III, IV and *P. sabit*	*P. galaxiae* and *P. sabit* sister clades (Teng 2015). *P. galaxiae* group II, III, IV strains confirmed by (McDonald et al. 2007, Lim et al. 2013)	7
None (multiple spp.)	*Pseudo-nitzschia pseudodelicatissima , P. cuspidata, P. plurisecta, P. fukuyoi, Fragilariopsis kurta, Fragilariopsis vanheurkii, Fragilariopsis kurguelensis, Fragilariopsis rhombica, Fragilariopsis cylindricus, Neodenticula seminae*	*P. pseudodelicatissima* confirmed by (Ajani et al. 2013, Lim et al. 2013). *P. cuspidata* confirmed by (Lundholm 2012, Ajani et al. 2013, Lim et al. 2013). *P fukuyoi* confirmed by (Lim et al. 2012, Lim et al. 2013). *P. pseudodelicatissima/cuspidata* complex confirmed by (Fernandes et al. 2014)	
None (multiple spp.)	*P. delicatissima, P. pseudodelicatissima, P. decipiens, P. galaxiae, P. fraudulenta, P. turdigula*	*P. delicatissima* identified as del 2 (Amato et al. 2007). No strain confirmation on other sequences.	2
Environmental 1 (this study)	None	None	1
Environmental 2 (this study)	*P. galaxiae, P. delicatissima*	Both public sequences identified from Ruggiero et al. 2015)	

973 **Table 4.** List of *Pseudo-nitzschia* species identified by SEM in from a subset of the genetically analysed sample set. Sample ID relates to the
974 CPR sample. The sequence analysis method indicates whether the samples were analysed using NGS 454 technology or clone library (CL)
975 sequencing technology. The number of raw reads generated from the HTS sequence analysis are indicated, where applicable.

CPR Sample	**method**	**HTS reads**	**Code**	***Pseudo-nitzschia* species found by SEM**	***Pseudo-nitzschia* species found by HTS**	***Pseudo-nitzschia* species found by CLS**
21VJ1	454	3178	Au02E	*P. heimii plus small undetermined species*	*P. fraudulenta*	N/A
21VJ5	CL		Au02E(2)	N/A	N/A	*P. fraudulenta*
139VJ1	454, CL	5001	Su08E	None (*Thalassiosira* spp. abundant)	*P. fraudulenta, P. multiseries*	*P. fraudulenta, P. multiseries*
146VJ5	454, CL	5902	Au08E	N/A	*P. multiseries, P. fraudulenta, P. subfraudulenta, P. galaxiae* II, III, IV/*P. sabit, P. abrensis/P. batesiana, P. subfraudulenta*	*P. multiseries, P. fraudulenta, P. pungens*
83VJ5	454	4506	Sp05E	N/A	*P. fraudulenta, P. multiseries,* Environmental 2	N/A
77VJ7	454	5116	Sp05NE	None	*P. fraudulenta*	N/A
21VJ41	454	3720	Au02C	*P. cuspidata, P. turgidula, P. heimii*	*P. fraudulenta*	N/A
132VJ17	454	2872	Sp08C	*P. turgidula*	*P. galaxiae* II, III, IV/*P. sabit, P. galaxiae* I/*P. sabit,*	N/A

					P. subfraudulenta, P. seriata, P. multiseries, P. hasleana/ kodamae, P. fraudulenta, P. delicatissima/P. arenysensis (2), Environmental groups 1, 3.	
146VJ37	454, CL	2613	Au08C	*P. turgidula, P. inflatula, P. fraudulenta*	*P. multiseries, P. galaxiae* II, III, IV/*.sabit, P. fraudulenta*	*P. fraudulenta*
83VJ41	454, CL	6459	Sp05C	*P. australis*	*P. multiseries, P. pungens, P. galaxiae* II, III, IV/*P.sabit*	*P. fraudulenta*
21VJ45	CL		Au02C(2)	*P. cuspidata, P. turgidula, P. australis*	N/A	*P. fraudulenta, P. multiseries*

976

977

978 **Table 5:** List of *Pseudo-nitzschia* species identified by SEM in from 2014. Sample ID relates to the CPR sample, locations shown in Fig.1.
979 Samples are listed following their longitudinal position.

Sample_id	Environment	Sample latitude	Sample longitude	month	year	*Pseudo-nitzschia* species found by SEM
272VJ-1	Eastern, Coastal	48.348	-124.135	8	2014	*P. fraudulenta, P. pungens, P. seriata, P. multiseries*
272VJ-5	Eastern, Coastal	48.517	-125.077	8	2014	*P. fraudulenta, P. pungens, P. seriata, P. multiseries*
272VJ-9	Eastern, coastal	48.73	-126.02	1	2014	*P. fraudulenta, P. pungens, P. seriata, P. multiseries*

273VJ-3	Central, Open	51.393	-136.688	2	2014	*P. inflatula, P. pseudodelicatissima, P. turgidula*
273VJ-11	Central, Open	51.848	-138.707	3	2014	*P. inflatula, P. pseudodelicatissima, P. turgidula*
272VJ-45	Central, Open	50.962	-134.648	9	2014	*P. inflatula, P. pseudodelicatissima, P. turgidula*
273VJ-39	Central, Open	53.067	-145.865	4	2014	*P. inflatula, P. pseudodelicatissima, P. turgidula*
273VJ-43	Central, Open	53.187	-146.96	4	2014	*P. inflatula, P. pseudodelicatissima, P. turgidula*

994

995 **Supplementary Material**

996 **Table A1:** General and *Pseudo-nitzschia* primers used in this study with references indicated in parentheses. Those highlighted in bold were
997 used to obtain the sequences presented in this study

998

999 **Table A2**: Identified HTS and clone-library derived environmental sequences. Acc no=Accession number.

1000

1001 **Table A3**: Alignment of Environmental sequences against publically available reference sequences.

1002

1003 **Fig. A1**: Maximum likelihood Phylogenetic analysis of CLS from a 430bp alignment of partial LSU fragment derived from D1-186F-D1-548R
1004 PCR amplification with- publically available LSU reference set of *Pseudo-nitzschia* sequences along with other diatom sequences as an
1005 outgroup. Grey boxes indicate clades that correspond to Fig. 2B. Asterix indicates sequences recovered together in Fig, 2C.

1006

1007 **Fig. A2**: Histogram of intraspecific and interspecific pairwise distances from publically available *Pseudo-nitzschia* species. Dark and light grey
1008 bars show intra- and inter-specific diversity respectively. Interspecific diversity overlaps with intraspecific diversity indicating a lack of
1009 boundary that could be useful to delineate species by genetic distances.

1010

1011 **Fig. A3**: LSU Maximum Likelihood (ML) phylogeny of Fig. 2, shown with genetic distances from a 439bp alignment of partial LSU fragment
1012 from public reference sequences and environmental *Pseudo-nitzschia* sequences. Clades without environmental sequences are collapsed for
1013 clarity.

Fig. A4. Time series of SST standardized anomalies for grid cells A-G in Fig. 4 that encompass the locations of CPR samples used in the molecular analyses. A = Au02C; B = Au02C(2) and Sp05C; C = Au08C; D = Sp08C; E = Sp05NE; F = Sp05E and Au08E; G = Au02E, Au02E(2) and Su08E. See Table 1 for details of sample locations. On each panel, asterisks indicate when the CPR samples were taken.